THE GEOMETRY OF
MENTAL MEASUREMENT

Sir GODFREY H. THOMSON

UNIVERSITY OF LONDON PRESS Ltd.

THE GEOMETRY OF
MENTAL MEASUREMENT

/c

THE GEOMETRY OF
MENTAL MEASUREMENT

by

SIR GODFREY H. THOMSON
D.C.L., D.SC., PH.D.
University of Edinburgh

UNIVERSITY OF LONDON PRESS LTD.
WARWICK SQUARE, LONDON, E.C. 4

Printed in Great Britain by
NEILL & CO. LTD., EDINBURGH

PREFACE

This little book had its origin in three Special University Lectures given in the Beveridge Hall of the Senate House, London University, in January 1953. Those lectures were addressed to students of psychology, and gave a geometrical approach to psychometric problems; but it is hoped that these pages may interest a wider audience, for many of the formulæ have been applied, or are applicable, in other sciences. The greater part of the book requires only a very meagre mathematical equipment from the reader, little more than the knowledge that the square on the hypotenuse of a right-angled triangle equals the sum of the squares on the sides, the meaning of the cosine and sine of an angle, and two simple trigonometrical formulæ. Other assumptions occur but are explained, and the later part of Section Eight is indicated as meant for a more advanced reader, and can be omitted.

What *is* demanded from the reader is a willingness to think spatially, even in terms of spaces of higher dimensions than three. Indeed, these pages may well form an introduction, for a budding mathematician, to the later more strenuous study of n-dimensional geometry.

GODFREY THOMSON

EDINBURGH
March 1953

CONTENTS

SECTION ONE

THE MODEL

THE primary purpose of this book is to describe a geometrical model from which can be deduced most of the formulæ used in the factorial analysis of human ability. Since many of these are, however, also used in other fields (for example the formulæ concerning selection, and partial regression co-efficients), it is hoped that the book may have a wider appeal.

The model is one in many dimensions, but it is best approached by dealing with its simpler forms first, in two and in three dimensions. Consider first the ordinary 'scattergram' illustrating the correlation between two variates, the one variate being measured along the x axis, the other along the y axis at right angles to the former, and each associated pair of values being represented by its point (x, y). If each variate is distributed normally, these points will vary in density over the diagram, being thickest at the position corresponding to the average of both variates and having elliptical density contours round that point. If, further, each variate is measured from its average, and in units of its own standard deviation, the contours of density of the points will be ellipses

$$x^2 - 2rxy + y^2 = \text{constant} \qquad (1)$$

and the major axis of these ellipses will be equally inclined to the co-ordinate axes. The correlation is measured by the quantity r (see Figure 1).

This quantity r can be calculated from the paired values of x and y by the formula

$$r = \frac{\text{Sum }(xy)}{\sqrt{\{\text{Sum }(x^2) \times \text{Sum }(y^2)\}}}$$

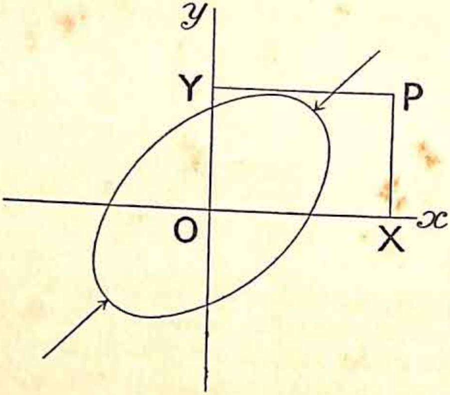

FIG. 1

where x and y are measured from their mean values, or by various forms of this formula. The standard deviations σ_x and σ_y are found from

$$n\sigma_x{}^2 = \text{Sum } (x^2)$$

and a corresponding formula for $\sigma_y{}^2$. If the standard deviations are taken as units in measuring x and y the previous formula becomes simply

$$r = \text{Sum } (xy)/n$$

x and y being now measured in standard units.

It is assumed that the reader has some acquaintance with what is meant by a normal distribution. It is a distribution in which values near the average are most frequent, and the frequency falls off as the average is departed from, in either direction. A distribution of binomial coefficients such as

$$1 \quad 4 \quad 6 \quad 4 \quad 1$$

or

$$1 \quad 6 \quad 15 \quad 20 \quad 15 \quad 6 \quad 1$$

is a distribution fulfilling these two requirements, but a normal distribution is a smooth continuous one, and may be looked

upon as a binomial distribution with a very large number of terms, very close together. Its actual formula, when measured from the middle, is

$$y = \frac{1}{\sigma\sqrt{(2\pi)}} e^{-x^2/(2\sigma^2)}$$

where e is the base of Naperian logarithms.

At this point it is desirable to say that throughout this book these two assumptions—that the distribution of each variate is normal and that it is measured in units of its own standard deviation—are made. Some of the formulæ proved by using the geometrical model have, however, a validity wider than for normally distributed variates, and all of them are approximately true when the departure from normal distribution is not great. As for measuring in units of standard deviation, this only means that if the variates are in practice measured in other units, it is necessary to change to the standard units before using the model to prove anything. Any formula can thereafter be changed back into any arbitrary units.

In the above scattergram outlined in Figure 1, if the two variates are scores in two mental tests, the points like P in the figure represent persons who get scores equal to their co-ordinates. Thus P represents a man or a child who scores OX and OY in the two tests. The person P, and every person whose point is in the north-east quadrant, is above average in both tests. Those in the south-west quadrant are below average in both. Those in the other two quadrants are above average in one, below average in the other, test.

Along each of the rectangular co-ordinate axes x and y there are really two coincident lines, one of them representing the variate (the test score), the other separating the people who do well in the other test from those who do badly. As long as the test-lines are at right angles to each other, these two lines remain coincident. But the change in the diagram about to be described will cause them to separate. We are about to compress the crowd of persons along the major axis until their contours of density are no longer ellipses, but circles.

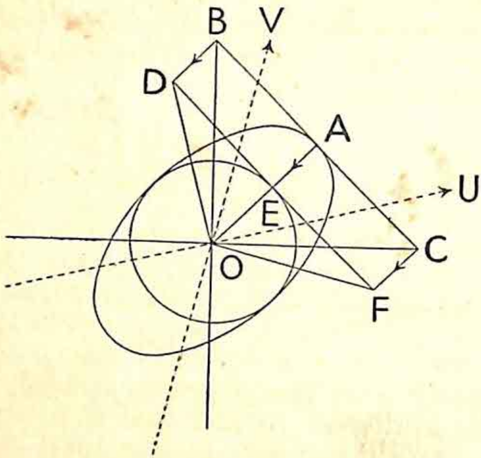

FIG. 2

Before doing so we note, what is easily proved from the formula $x^2-2rxy+y^2=$constant, that the major and minor axes are in the ratio

$$\frac{\text{minor axis}}{\text{major axis}}=\frac{\sqrt{(1-r)}}{\sqrt{(1+r)}} \qquad (2)$$

In Figure 2 we are therefore going to move a line BAC (at right angles to the major axis) parallel to itself until it reaches DEF, where OE is now equal to half the minor axis. Imagine this line BAC pushing the crowd of persons before it until, when it reaches DEF, their contours of density have changed from ellipses into circles—for a similar compression towards O must be envisaged from the other end of the major axis also. This double compression (which must be understood as going on suitably for every elliptical contour) has moved the line OB to OD, and OC to OF. We then have, if we take $OA^2=1+r$ and $OE^2=1-r$,

$$OF^2=OE^2+EF^2=OE^2+OA^2=2$$

and

$$\cos DOF=2\cos^2 EOF-1=2\frac{1-r}{2}-1=-r$$

These lines OD and OF are still the lines which separate 'above average' from 'below average' persons. All the persons who were in the quadrant BOC, and therefore good in both tests, are now in the wider sector DOF. We still want their scores in the tests to be represented by the vertical projection of their points on to the test-lines. It is therefore evident that the two test-lines must be at right angles to OD and OF respectively, in the dotted positions shown in Figure 2, and the angle between those dotted lines (the test-lines) is the supplement of DOF and its cosine is r, the correlation coefficient.

We have changed from a diagram in which the existence of correlation is shown by the ellipticity of the density contours of points representing persons, to a diagram in which the crowd of points representing persons is circular in contour, and in which the correlation between two tests is shown by their lines being, not at right angles, but at an angle whose cosine is r.

It will be noticed that by compressing the crowd in the way we have done, namely inward along the major axis, we have altered the scale of the diagram, so that all standard deviations have sunk to below unity. We ought really to have made our change from ellipse to circle by partly compressing the major axis and partly stretching the minor axis: but that would have unduly complicated Figure 2, and we can attain the same end result by imagining the circle expanded until the scale is once more unity, a procedure which will not change any angles and in particular will leave cos UOV $= r$. In the diagram then reached, each test-line can be looked upon as a 'vector'. A vector is a direction with a weight or strength attached to it. In our model the weight of each is the standard deviation, in the full test-line unity.

If the number of persons above average in both tests be a, the same number will, under our assumption of normal distribution, be below average in both; and b will be above in one and below in the other, with an equal number b below in the one and above in the other. Those above average in both are, in our Figure 2, all those in the sector DOF: and in this

circular crowd it is evident that the angle DOF is $\frac{a}{a+b}\pi$, while the supplement to DOF is $\frac{b}{a+b}\pi$. But the angle UOV between the dotted test-lines is the supplement of DOF, and its cosine is r. So we have Sheppard's formula

$$r=\cos\frac{b}{a+b}\pi$$

a formula only applicable when the cuts forming a fourfold table are actually through the means of the two variates.

If a third test (variate) is added to these two, we must look up the angles whose cosines equal its correlation coefficients with them, and place the line representing the new test so as to make these angles with the existing lines. This third line will not, in general, be in their plane but will project into a third dimension. The points representing persons are now a swarm with spherical contours of density. Any straight line through the origin O (where are situated the persons average in all three tests) will represent a test which is a weighted combination of the first three.

A fourth test will in general project into a fourth dimension, and a model of n tests will be in n dimensions. The crowd of person-points will still be 'spherical' in contour.

Along any straight line through the crowd of points their density distribution is normal, with unit standard deviation.

SECTION TWO

SELECTION

WITH the aid of a three-dimensional model let us first examine the changes in correlation coefficients produced by a selection which reduces the scatter of a variate. In Figure 3 let OA, OB and OC be the lines representing three tests, and let us call the angles AOC, BOC and AOB, α, β and γ. Suppose that in test (variate) 3 a selection of persons is made such that their standard deviation on scores in that test is reduced from unity to p_3. What is the effect on the correlations between the three tests? Let us suppose in the first place that the average score in test 3 is not altered by the selection.

Let ACB be a plane at right angles to OC, and take OC as unity. Then the selection will result in the plane ACB being replaced by one parallel to it in the position A′C′B′ where $OC' = p_3$. A corresponding movement towards O must be imagined in the upper part of Figure 3.

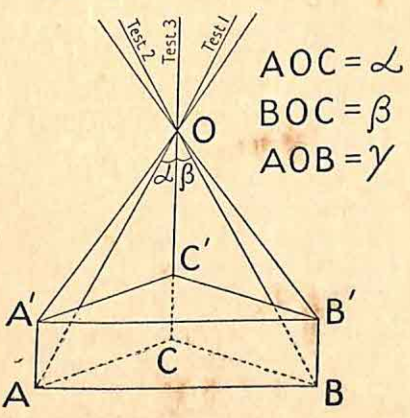

FIG. 3

Before these selections and resulting changes in the diagram, the correlation coefficients between the tests were $r_{13}=\cos\alpha$, $r_{23}=\cos\beta$, and $r_{12}=\cos\gamma$. The new correlation coefficients are the cosines of the angles A′OC′ or α', B′OC′ or β', and A′OB′ or γ'. Let us write $p^2+q^2=1$ and find the values, in the figure, of p and q for each test. We have

$$p_3=\text{OC}',\qquad p_1=\frac{\text{OA}'}{\text{OA}},\qquad p_2=\frac{\text{OB}'}{\text{OB}}$$

Further,

$$\text{OA}^2-\text{OA}'^2=\text{OB}^2-\text{OB}'^2=\text{OC}^2-\text{OC}'^2=q_3{}^2 \tag{3}$$

Also

$$\left.\begin{aligned} q_1{}^2=1-p_1{}^2=\frac{\text{OA}^2-\text{OA}'^2}{\text{OA}^2}=\frac{q_3{}^2}{\text{OA}^2}&=q_3{}^2\cos^2\alpha \\ \text{and similarly}\qquad q_2{}^2&=q_3{}^2\cos^2\beta \end{aligned}\right\} \tag{4}$$

That is, $q_1=q_3r_{13}$ and $q_2=q_3r_{23}$. It is test 3 which has been directly selected. The other two tests have had their scatter indirectly reduced as a result, and their q's are got from q_3 by multiplying it by their correlation coefficients with the directly selected test.

The correlation between the directly selected test 3 and test 1 is

$$\cos \text{A}'\text{OC}'=\frac{\text{OC}'}{\text{OA}'}=\frac{p_3}{p_1\text{OA}}=\frac{p_3}{p_1}\cos\alpha$$

i.e. the new $$r'_{13}=\frac{p_3}{p_1}r_{13}$$

Similarly $$r'_{23}=\frac{p_3}{p_2}r_{23}$$

The correlation between the two indirectly selected tests 1 and 2 is a little more complicated. We have

$$\text{AB}^2=\text{OA}^2+\text{OB}^2-2\text{OA}.\text{OB}\cos\gamma$$

and $$\text{A}'\text{B}'^2=\text{OA}'^2+\text{OB}'^2-2\text{OA}'.\text{OB}'\cos\gamma'$$

whence, using equations (3) and (4), we get, since $AB = A'B'$,

$$\cos \gamma' = \frac{OA . OB \cos \gamma - q_3^2}{OA' . OB'} = \frac{\cos \gamma - \dfrac{q_3}{OA} \cdot \dfrac{q_3}{OB}}{p_1 p_2} = \frac{\cos \gamma - q_1 q_2}{p_1 p_2}$$

i.e. the new

$$r'_{12} = \frac{r_{12} - q_1 q_2}{p_1 p_2} \tag{5}$$

This more general formula (5) includes as a special case the formula for the partial correlation $r_{12.3}$, where the selected test 3 is restricted to one value and its standard deviation is zero. That is, $p_3 = 0$ and $q_3 = 1$. Then

$$\left.\begin{matrix} q_1 = \cos \alpha \\ q_2 = \cos \beta \end{matrix}\right\} \text{ from equation (4)}$$

and formula (5) becomes

$$r_{12.3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{\{(1 - r^2_{13})(1 - r^2_{23})\}}} \tag{6}$$

We have so far assumed that the selection of test 3 was such that its average was unchanged. In our Figure 3 we have not altered the position of O. But a study of Figure 3 makes it evident that if instead of moving the plane ACB up towards O we had moved O down by the same amount towards ACB, to a new position O′, we would not have changed anything essential, and could with practically the same trigonometry have arrived at exactly the same formula (5). Or we might have moved ACB a little up and O a little down, until the distance between them was again p_3. In short, provided the distributions remain normal, a change in the mean caused by selection still allows formulæ (5) and (6) to be used. And in fact, even if the distributions are not quite normal, these formulæ have a kind of average validity.

SECTION THREE

PROJECTION TO LINES, PLANES AND SPACES

THIS section is devoted to the statement of a very simple principle of geometry which will be frequently used in the pages which follow. Most readers will probably be already familiar with it, but a few may not, especially in higher dimensions.

By the projection of a point A on to a line XY is meant the point C at the foot of the perpendicular from A on to XY (see Figure 4). By the projection of a distance AB on to the line is meant the distance CD in the figure.

$$CD = AB \cos \theta = AB \sin \phi$$

The lines AB and XY need not be in the same plane. If not, the angle θ is the angle between AB and a line parallel to XY and cutting AB.

Clearly the projection of a chain of straight lines, beginning at A and ending at Z, on to a line XY, is the same as the projection of AZ on to that line. The lines forming the chain can trespass into a space of many dimensions, provided each one begins where the preceding one ended. A point can also be projected on to a plane, as in Figure 4, where DP is perpendicular to the plane of the table-top. If P is then projected on to any line OE in the plane, the point Q thus reached is the same as the direct projection of D on to the line OE.

This is evident from the fact that the plane DQP is at right angles to the plane of the table-top and therefore to OE, hence OQD is a right angle. Or it can be proved more laboriously thus. Take any point O in OE and join it to P and to D. Then by our construction the angles DPO, DPQ and PQO are right angles.

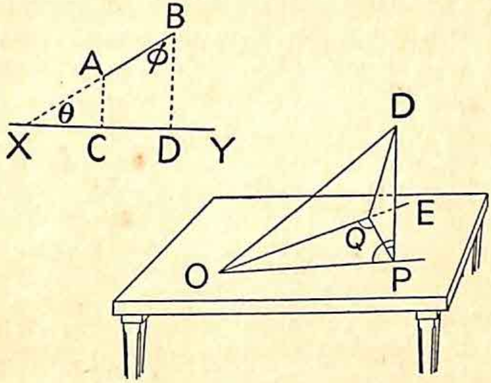

FIG. 4

Therefore $$OD^2 = DP^2 + OP^2$$

and $$OP^2 = OQ^2 + QP^2$$

i.e. $$OD^2 = DP^2 + OQ^2 + QP^2$$

But $$DP^2 + QP^2 = DQ^2$$

therefore $$OD^2 = DQ^2 + OQ^2$$

and so OQD must be a right angle

The principle here seen in three dimensions is also true in many dimensions. If instead of the table-top we imagine a space of n dimensions, and a point D outside that space, then D can be projected on to that n-space at a point P (for a perpendicular from outside a space on to the space hits it at a point and goes through it at once: the space has no thickness in that direction). If then P is projected on to a subspace of m dimensions inside the n-space ($m < n$), it will hit that subspace at the same point Q as though D had been projected directly on to the subspace. For $n = 3$ and $m = 2$ (and 1) this is illustrated in Figure 5, where by a kind of superperspective four dimensions are portrayed on the paper.

A 3-space is defined by the lines OU, OV and OW (which

may be three test-lines). D is not in that 3-space but outside it, and DP is a perpendicular to the space. From D also a perpendicular DQ is drawn to the plane defined by OU and OV. Then join PQ.

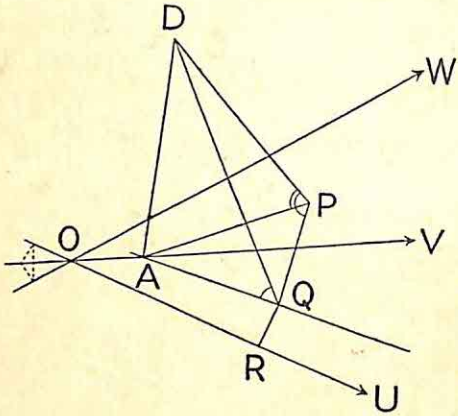

Fig. 5

AQ is any line through Q in the plane UOV, and A is then joined to P and D. It has now to be shown that PQ is at right angles to the plane UOV. (The angles don't look like right angles, but that is because of the superspective. In ordinary perspective right angles don't always or indeed usually look like right angles.)

Because DP is at right angles to the whole 3-space OUVW, the angles DPA and DPQ are right angles: and because DQ is at right angles to the plane UOV, DQA and DQR are right angles.

Therefore $$DA^2 = DQ^2 + QA^2$$
and also $$= DP^2 + PA^2$$
Further, $$DQ^2 = DP^2 + PQ^2$$
so $$DP^2 + PA^2 = (DP^2 + PQ^2) + QA^2$$
or $$PA^2 = PQ^2 + QA^2$$
so PQA is a right angle

And since AQ was any line in the plane UOV through Q, PQ is at right angles to that plane. The same point Q is reached whether D is projected directly on to the plane UOV, or first on to the space OUVW and thence on to UOV.

Similar reasoning shows that R, the projection of Q on to OU ($m=1$), is also the projection of D and of P.

SECTION FOUR

HIGHER SPACES

SPACES of more dimensions than three are of course only a manner of speech. There is no suggestion that such spaces actually exist in the sense our familiar three-dimensional space exists for us who live in it. But with a little care, and some acceptance of new features, many of the geometrical laws we know in our familiar space are still true in higher dimensions, and it is convenient to use the familiar terms. For example, a sphere is a surface everywhere the same distance from a point called its centre, and its equation is

$$x^2 + y^2 + z^2 = \text{radius}^2$$

In a higher space, say of five dimensions, we still give the name sphere, or hypersphere, to a surface whose equation is

$$x^2 + y^2 + z^2 + v^2 + w^2 = \text{radius}^2$$

Some things which happen in higher space are rather startling to the layman. For instance, in a seven-space one can have a subspace of four dimensions, and another of three dimensions, which are completely orthogonal to one another, so that every line in the one space is at right angles to every line in the other. Such completely orthogonal spaces have only one point in common with each other. The analogue in our ordinary space is a line (of one dimension) perpendicular to a plane (of two dimensions). The sum of the dimensions of two such orthogonal spaces cannot be greater than the number of dimensions of the space containing them.

An instance at first sight apparently contradicting this is two planes at right angles to one another in our familiar three-space, like a drawing-board standing on edge on a

table-top. They have a whole line in common; and the sum of their dimensions is four, greater than three. But these two planes are not *completely* orthogonal. Many lines can be drawn on the drawing-board not at right angles to the table-top.

As was earlier said, a perpendicular from a point outside a space to that space goes clean through it at a point. Further, a line not in the space can be projected on to it, giving there a line. Once these ideas become familiar they cause no trouble: but to a newcomer they are strange.

SECTION FIVE

THE COSINE LAW

IN a three-space we can, given a set of three co-ordinate axes, define a point P (see Figure 6) by its three ordinates, x, y and z. A line from the origin O through P we can define by the cosines of the three angles it makes with the co-ordinate axes. These are called its direction cosines. If the distance OP is taken as unity, then the co-ordinates $x = \text{OA}$, $y = \text{AB}$ and $z = \text{BP}$ are also the direction cosines of the line OP. And the sum of their squares is unity. For $\text{OA}^2 + \text{AB}^2 = \text{OB}^2$, and then $\text{OB}^2 + \text{BP}^2 = \text{OP}^2 = \text{unity}$. In a higher space we can again define a direction by its direction cosines, whose squares there too sum to unity.

If we have another line OQ (OQ = unity) with its direction cosines OC, CD and DQ, then the cosine of the angle POQ is equal to

$$\text{OA} \times \text{OC} + \text{AB} \times \text{CD} + \text{BP} \times \text{DQ}$$

We call this 'forming the inner product of the direction cosines'. Each is multiplied by its opposite number, and the products are summed.

The proof is almost self-evident. If OP is projected on to OQ (see Figure 6), then since OP is unity, OR is the cosine of POQ. Instead of projecting OP itself on to OQ, project the chain of lines OA, AB and

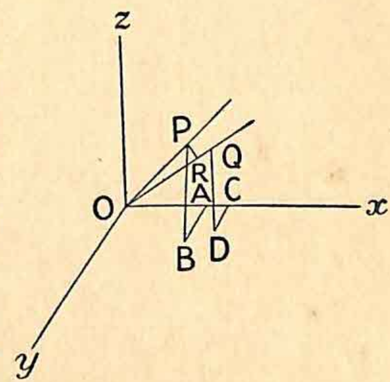

FIG. 6

BP by multiplying each by the appropriate direction cosine of OQ. Thus

$$OR = OA \times OC + AB \times CD + BP \times DQ$$

If in another notation we call the angles OP makes with the axes α, β and γ and those which OQ makes α', β' and γ', then

$$\cos POQ = \cos \alpha \cos \alpha' + \cos \beta \cos \beta' + \cos \gamma \cos \gamma'$$

It is clear that the above proof would still apply in any number of dimensions. The inner product of their direction cosines is the cosine of the angle between two lines. In two dimensions this is the familiar formula

$$\cos (A - B) = \cos A \cos B + \sin A \sin B$$

SECTION SIX

CENTROID-FACTORS USING UNIT COMMUNALITIES

In factorial analysis the matrix of intercorrelations of a number of variates is often first 'factorised' by a process known variously as 'centroid analysis' and 'simple summation'. The latter name reflects part of the arithmetical procedure,

$$\begin{array}{ccc}
1{\cdot}0 & {\cdot}7 & {\cdot}3 \\
{\cdot}7 & 1{\cdot}0 & {\cdot}8 \\
{\cdot}3 & {\cdot}8 & 1{\cdot}0 \\
\hline
2{\cdot}0 & 2{\cdot}5 & 2{\cdot}1 \quad = 6{\cdot}6 = 2{\cdot}569^2 \\
{\cdot}779 & {\cdot}973 & {\cdot}817 \\
l_1 & l_2 & l_3
\end{array}$$

the former reflects the fact that the variates are being treated as 'vectors' in a space equal in dimensions to their number. If the matrix of correlations is completed by inserting unity in each diagonal cell ('unit' communalities), the arithmetical process for extracting the first 'factor' is illustrated above, for three tests or variates only. The columns are summed, giving the values 2·0, 2·5 and 2·1. These sum to 6·6, whose square root is 2·569. The column sums, divided by this last quantity, give what are called the loadings, or the saturations, of each test with this first factor. They are the correlation coefficients of each variate with this factor, itself a hypothetical variate.

In geometrical language each of the intercorrelations is the cosine of an angle between the lines representing the two variates or tests, and the loadings are the cosines of the angles each test-line makes with their resultant or centroid—hence

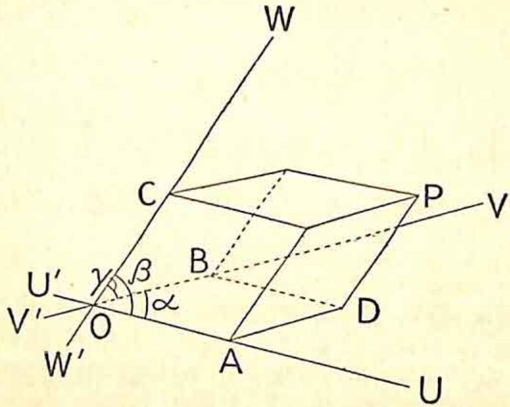

FIG. 7

the name centroid method. Our first purpose is to show that the above arithmetical process does correspond with this geometrical picture. We shall deal first with the case of three tests. Let their intercorrelations be

$$\left.\begin{matrix} 1 & \cos\alpha & \cos\beta \\ \cos\alpha & 1 & \cos\gamma \\ \cos\beta & \cos\gamma & 1 \end{matrix}\right\} \qquad (7)$$

and let these three tests be represented in Figure 7 by the three lines U′OU, V′OV and W′OW with the angles α, β and γ between the pairs. (Here, as in most later diagrams, the figure is drawn on the positive halves of the test-lines, and acute angles are used between the latter, corresponding to positive correlation coefficients. But these restrictions are only for convenience in drawing the figures.)

The resultant of three equal forces along these lines lies along the diagonal OP of the parallelopiped shown in the figure, with sides each equal to unity. Let the angle with OP made by each test-line be ϕ, θ, ω, so that the correlations of the tests with this resultant or centroid are $\cos\phi$, $\cos\theta$ and $\cos\omega$.

Now project OP on to OU, first directly, and secondly by

projecting the chain of lines OA, AD and DP. This gives us the equation

$$\text{OP} \cos \phi = \text{OA} + \text{AD} \cos \alpha + \text{DP} \cos \beta$$

$$\left.\begin{aligned} &= 1 + \cos \alpha + \cos \beta \\ &\text{and similarly} \\ \text{OP} \cos \theta &= \cos \alpha + 1 + \cos \gamma \\ \text{OP} \cos \omega &= \cos \beta + \cos \gamma + 1 \end{aligned}\right\} \qquad (8)$$

From these we see that

$$\cos \phi + \cos \theta + \cos \omega = \frac{\text{Sum of all the items in the matrix (7)}}{\text{OP}}$$

But $\cos \phi + \cos \theta + \cos \omega = \text{OP}$, for this expression is the projection on to OP of the chain of lines OA, AD, DP. We have therefore that

$$\text{OP}^2 = \text{Sum of all the items in the matrix}$$
$$(= 6{\cdot}6 \text{ in our arithmetical example})$$

and from the first equation of (8),

$$\cos \phi = \frac{1 + \cos \alpha + \cos \beta}{\text{OP}}$$

$$= \frac{\text{Sum of the first column of the matrix}}{\text{Square root of the total sum}}$$

which is exactly the procedure of simple summation. The quantity $\cos \phi$ is the correlation coefficient of test 1 with the centroid, the loading or saturation.

Our diagram is in three dimensions only, but clearly the proof is general in any number of dimensions.

Let us now return to the arithmetical example. The loadings of each test with the first centroid factor having been obtained (let us call them l_1, l_2, l_3), each entry in the original matrix of correlations is reduced by the part explained by that first factor. Thus r_{23} is reduced to $r_{23} - l_2 l_3$ (in our example,

$\cdot 8$ is reduced to $\cdot 8 - \cdot 973 \times \cdot 817 = \cdot 005$). The units in the diagonal cells are reduced by the square of the corresponding loading; the first one becomes, *e.g.*, $1{\cdot}000 - \cdot 779^2 = \cdot 394$. Thus we arrive at the matrix of 'residues':

$$\begin{bmatrix} 1 - l_1^2 & r_{12} - l_1 l_2 & r_{13} - l_1 l_3 \\ r_{12} - l_1 l_2 & 1 - l_2^2 & r_{23} - l_2 l_3 \\ r_{13} - l_1 l_3 & r_{23} - l_2 l_3 & 1 - l_3^2 \end{bmatrix} = \begin{bmatrix} \cdot 394 & -\cdot 058 & -\cdot 336 \\ -\cdot 058 & \cdot 053 & \cdot 005 \\ -\cdot 336 & \cdot 005 & \cdot 332 \end{bmatrix}$$

It will be seen that each column sums to zero.

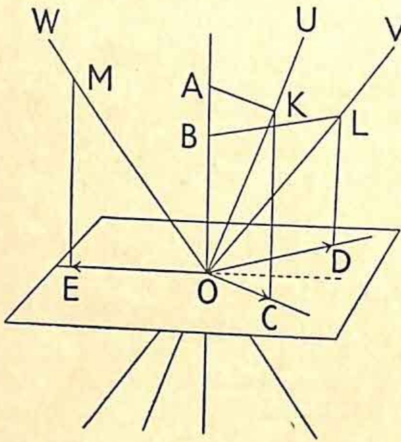

FIG. 8

We must now see what, in the geometrical model, corresponds to this arithmetical procedure, and to the fact that these columns add up to zero. Let us deal with the case of three tests first, to enable ordinary diagrams to be made (Figure 8). OU, OV and OW are three test-lines, turned (for the diagram) so that their centroid OBA is pointing upward. The lengths OK, OL and OM are each equal to unity, the entry in the diagonal cells. Their loadings with the centroid, being their correlation coefficients with it, are the cosines of the angles between each and OA, and since the lengths OL, etc. are unity, these cosines are the distances OA for the first test, OB for

the second, and a similar projection (not shown) of OM on to OBA. $OA = l_1$ and $OB = l_2$.

Removing the influence of the centroid 'factor' means projecting the test-lines as shown on to a space of one less dimension, here in this simple case on to a plane through O at right angles to the centroid, and dealing only with their components in that plane. These components have weights equal to the lengths OC, OD and OE, and with these weights they are in equilibrium, for that is the nature of a centroid, that there is no sideways pull. The angles between the lines on the plane can be found by the formula for partial correlation (equation (6), page 17). Thus the cosine of DOC is

$$\cos \mathrm{DOC} = \frac{\cos \mathrm{KOL} - \cos \mathrm{KOA} . \cos \mathrm{LOB}}{\sqrt{\{(1 - \cos^2 \mathrm{KOA})(1 - \cos^2 \mathrm{LOB})\}}}$$

$$= \frac{r_{12} - l_1 l_2}{k_1 k_2} \quad \text{if we write } k^2 = 1 - l^2$$

Similarly formulæ are obtainable for cos DOE and cos EOC. Now since the three 'forces' (as we may call them to recall the laws of combination of forces) OC, OD and OE are in equilibrium, the projections of OD and OE on to the line OC must just balance OC. That is, the sum of

$$\mathrm{OC} \quad \text{or} \quad k_1$$

$$\mathrm{OD} \cos \mathrm{DOC} \quad \text{or} \quad k_2 \times \frac{r_{12} - l_1 l_2}{k_1 k_2}$$

and

$$\mathrm{OE} \cos \mathrm{EOC} \quad \text{or} \quad k_3 \times \frac{r_{13} - l_1 l_3}{k_1 k_3}$$

must be zero. If we multiply each of these quantities by k_1 their sum will still be zero: that is,

$$\begin{array}{ll} & k_1^2 \\ & r_{12} - l_1 l_2 \\ & r_{13} - l_1 l_3 \\ \hline \text{sum to} & \text{zero} \\ \hline \end{array}$$

But this is exactly the first column of the matrix of residues (page 28).

This argument also holds in a space of many dimensions, say n dimensions containing n test-lines. If unities are used in the diagonal cells of the $n \times n$ matrix of correlation coefficients, the centroid factor is also in that n-space, and by its direction defines one of n rectangular co-ordinates. At right angles to the centroid, through the origin, is an $(n-1)$-space in which the n remainders or components of the test-lines exist.

Any two of the original test-lines, say tests 1 and 2, and the centroid, define a three-space, and the two 'remainders' are in a plane at right angles to the centroid, a plane which forms part of the $(n-1)$-space. To this three-space the previous argument applies unchanged, and the cosines of the angles between all the 'remainders' are as before, though these remainders are not now in a two-dimensional plane. They are, however, in equilibrium (a centroid being what it is), and when projected on to one of their number, say test-remainder 1, and multiplied by k_1, they give the numbers in the first column of the table of residues, which therefore add up to zero. Since these 'remainders' in the $(n-1)$-space are in equilibrium there is no question of finding a centroid of them as they stand, to be the second 'factor'. This difficulty is overcome by temporarily changing the sign, *i.e.* the direction, of some of them so that they once more form a pencil: and to make the second factor as effective as possible, directions are changed until the pencil is as narrow and compact as possible. In the arithmetical form this corresponds to changing the signs of some rows (and the corresponding columns) in the matrix of residues, until the sum-total of all the cells is as large as possible.

In the simple case of our three-dimensional drawing in Figure 8, where the $(n-1)$-space is a plane, it is OE which would be reversed to get a good narrow pencil on the plane.

The arithmetical procedure to find loadings on the second factor is then exactly the same as that for the first factor. The process then proceeds to subspaces of $(n-2)$, $(n-3)$ . . . dimensions until n orthogonal factors have replaced the oblique test-lines.

SECTION SEVEN

CENTROID-FACTORS USING MINIMUM COMMUNALITIES

THE preceding section dealt with the case where unit entries are made in the diagonal cells of the matrix of correlation coefficients between the tests. The whole process is then conducted in the test-space, *i.e.* the n-space defined by the n test-lines. It is, however, the common practice to insert fractions called communalities in those diagonal cells, with the hope of thereby abbreviating the process of extracting centroids, since the diagonal cells are exhausted sooner, and by a suitable choice of the inserted communalities it can be arranged that the other cells also vanish then: or at least that they become small enough to be disregarded. We shall assume, in our geometrical considerations, that they actually vanish. It does not concern us here to ask how these communalities which have this result are discovered: various ways of first guessing them exist, and the guesses are afterwards refined and improved. The communalities are made as small as possible without causing imaginary quantities containing $\sqrt{(-1)}$ to arise later in the arithmetical work.

Now it is obvious that in general a space defined by n test-lines cannot be defined by a smaller number, say c, of orthogonal centroids, and indeed it will appear that the c-space of the centroids (called 'common' factors) is not in the n-space of the tests at all.

Further, these c common factors do not account for the whole of each test but only for the communality. The balance to make up unity in each diagonal cell is still unaccounted for, and n so-called 'specific' axes are added to the c centroid axes to do this. Each of these specific axes is at right angles to all the test-lines except one. Since the communalities were

made as small as possible, the specifics were thereby made as large as possible.

We then have a space of $n+c$ dimensions divisible into a common-factor-space of c dimensions and a specific factor space of n dimensions. The test-space, also of n dimensions, is part of the $n+c$ space, but it neither coincides with the specific-factor-space nor contains the common-factor-space.

All this can be conveniently illustrated by the case of $n=2$. (Of course two tests would never be analysed into factors—we are merely using the simplest possible illustration.) In that case the matrix of correlations is

$$\begin{array}{cc} 1 & r_{12} \\ r_{12} & 1 \end{array}$$

If, however, we insert r_{12} into each diagonal cell in place of unity, the arithmetical process described on page 25 gives for loadings with the first centroid factor the values $l_1=l_2=\sqrt{r_{12}}$, thus:

$$\begin{array}{lcc} & r_{12} & r_{12} \\ & r_{12} & r_{12} \\ \hline & 2r_{12} & 2r_{12}=4r_{12}=(2\sqrt{r_{12}})^2 \\ \text{loadings} & \sqrt{r_{12}} & \sqrt{r_{12}} \end{array}$$

The residues left after removing the influence of the first factor are

$$\begin{array}{cc} r_{12}-(\sqrt{r_{12}})^2 & r_{12}-(\sqrt{r_{12}})^2 \\ r_{12}-(\sqrt{r_{12}})^2 & r_{12}-(\sqrt{r_{12}})^2 \end{array}$$

that is, zero: and the process stops after one common factor is extracted.

In Figure 9 the two tests are represented by the test-lines OU and OV, with an angle UOV whose cosine is r_{12}. The sole common factor is represented by the line OG, the angles UOG and VOG are equal, and the cosine of each is $\sqrt{r_{12}}$. Since two such angles cannot, when added, come exactly to the

angle whose cosine is r_{12}, the line OG does not lie in the plane UOV.

(In passing, the reader will no doubt remark that OG, not being in the plane of OU and OV, is not *their* centroid. Of that, more below.)

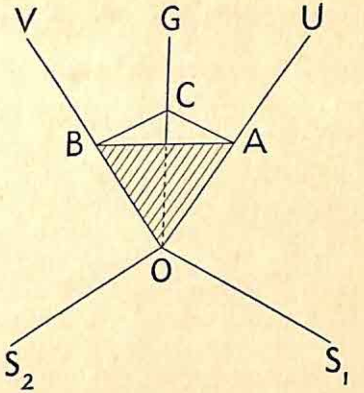

FIG. 9

The common-factor-space here is of one dimension only, and is clearly not in the test-space, which is the plane UOV. Two more axes must be added to OG to enable points in the test-space to be described (by three co-ordinates), and these, if they have to be specific and not common factors, have to be OS_1 perpendicular to the plane of GOV, and OS_2 perpendicular to the plane of GOU. The lines OG, OU and OS_1 are all in one plane, and the lines OG, OV and OS_2 all in another, perpendicular to it. (The points ABC and the cross-hatching are merely added to make the drawing look solid.)

In the more general case of n tests, when communalities are used, each test-line can be looked upon as having $c+1$ co-ordinates; c in the common-factor-space and 1 along the specific axis belonged to that test. With the c communal co-ordinates it makes angles whose cosines are its loadings with the successive factors obtained by the simple summation process, which is carried out in exactly the same way here as on page 25, except that communalities have replaced the units in the diagonal cells. If, following custom, we call the communalities in a 'battery' of three tests h_1^2, h_2^2 and h_3^2, the process, for the first factor, is

$$\begin{array}{lll}
h_1^2 & r_{12} & r_{13} \\
r_{12} & h_2^2 & r_{23} \\
r_{13} & r_{23} & h_3^2 \\
\hline
h_1^2+r_{12}+r_{13} & h_2^2+r_{12}+r_{23} & h_3^2+r_{13}+r_{23}=\mathrm{T}=t^2
\end{array}$$

C

and the loadings are the column sums divided by

$$t=\sqrt{(h_1^2+h_2^2+h_3^2+2r_{12}+2r_{13}+2r_{23})}$$

Now this first factor, though not the centroid of the test-lines, is the centroid of their communal components in the common-factor-space, as we must shortly show: but first there are two preliminary points to clear up—the geometrical meaning of h the square root of a communality, and the size of the angles between the communal components of the test-lines in the common-factor-space. We shall show that h_i is the projection on the c.f.s. (a convenient abbreviation) of unit distance along the test-line of test i; and that, if ϕ be the angle between test-lines i and j, the angle θ between h_i and h_j has as cosine the value

$$\cos\theta=\frac{\cos\phi}{h_ih_j}=\frac{r_{ij}}{h_ih_j} \qquad (9)$$

In Figure 9 the c.f.s. was of one dimension only. In that figure, if OA is unit distance along OU, then OC is the loading of test U on the sole common factor OG: and OC is also the projection of unit distance of OU on to the c.f.s. Here (as can be seen from the algebraic work of page 32) $h_1^2=r_{12}$ and $h_1=\sqrt{r_{12}}$ the cosine of angle GOU, so that $OC=h_1$.

In the more general case, with numerous tests, it is unlikely that communalities can be found to reduce the number of centroids to one. The common-factor-space will have, say, c dimensions. A test-line OU is not in this c-space, but sticks out into a further dimension. We can describe OU by means of $c+1$ axes, namely the c factor axes in the c.f.s., and its specific axis; and its direction cosines (see Section Five, page 23) are the projection of unit distance like OA on to these $c+1$ axes. As with all sets of direction cosines, their squares sum to unity. Let us make a diagram of the case where $c=2$, that is, there are two common factors. These two common factors, and the specific of test i, define a three-space, shown in Figure 10. Here OU is the test-line (in its positive half, O being the average point) and OA unit distance along it. Two common factors lie along OX and OY, and OS_i is the specific

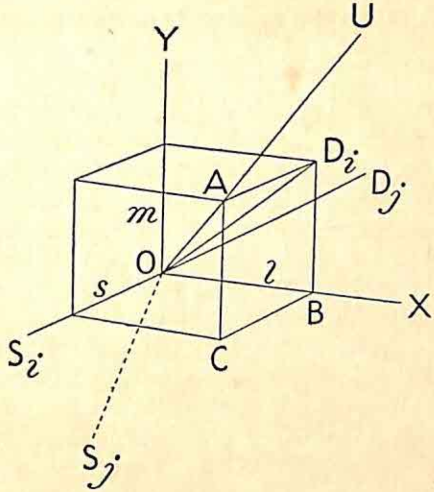

FIG. 10

of OU, at right angles to the plane XOY. The co-ordinates of A on these three axes are OB, BC and CA, the loadings of test i on the two common factors and its specific, or, in other words, the direction cosines of OU. The distance OA projected on to the plane XOY gives the distance OD_i whose co-ordinates on that plane are OB and BD_i equal to two of the co-ordinates of A, the two common factor loadings of the test OU. If we call the three loadings l, m and s (see the figure), then

$$l^2+m^2+s^2=1$$

and $$OD_i^{\,2}=l^2+m^2=h_i^{\,2}$$

for the arithmetical procedure, which consisted in subtracting from the communality first l^2 and then m^2, by hypothesis has exhausted the communality and reduced it to zero. So $OD_i=h_i$. The quantities h for each test are the projections on to the common-factor-space of unit distance along each test-line. Clearly this, shown here only for the case of two common factors, is general for c common factors. In the $c+1$ space the loadings of OA still, when squared, sum to

unity: and its projection on to the common-factor-space is the square root of the sum of squares of the common factor loadings.

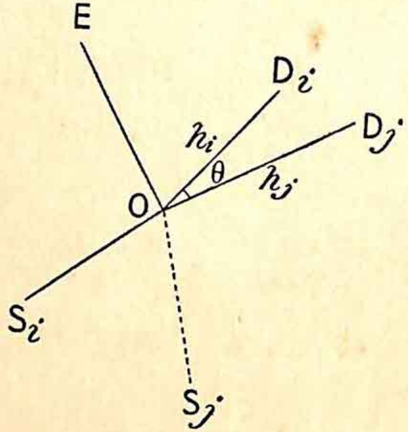

FIG. 11

Next consider the size of the angle θ between two lines like OD, the projections of two test-lines OU and OV. Two such are shown in Figure 10, OD_i and OD_j, and the specific axis OS_j of the new test is indicated by a dotted line. OV, the line of the new test itself, is not shown. It is in a different three-space, S_jOXY. The whole diagram is now in a four-space, and we want to find the cosine of the angle $\theta = D_iOD_j$.

In Figure 11 we have extracted from Figure 10 the essential lines for this purpose and have added a line OE in the common-factor-space at right angles to OD_j. Using as four rectangular co-ordinate axes the four lines OD_j, OE, OS_i, OS_j, we have, for the direction cosines of the two test-lines,

for OU	$h_i \cos \theta$	$h_i \sin \theta$	$\sqrt{(1-h_i^2)}$	0
for OV	h_j	0	0	$\sqrt{(1-h_j^2)}$

The inner product (see page 23) of these four direction cosines gives us the value of $\cos \phi$ (between the test-lines),

$$\cos \phi = h_i h_j \cos \theta$$

whence
$$\cos \theta = \cos \phi / h_i h_j = r_{ij} / h_i h_j$$

We can now proceed to the task of showing that the successive factors found by the process of simple summation using communalities are the centroids of the communal components of the tests. Take first the case of two common factors (Figure 12), and, for convenience in drawing the diagram,

three tests—but the argument which follows would apply to any number.

Take as three co-ordinate axes OS_1 (the specific of test 1), OL along the line of h_1 in the c.f.s., and a line OM in the c.f.s. at right angles to h_1. The figure has been decorated with a door, a window, and a picture to make it look solid (Figure 12).

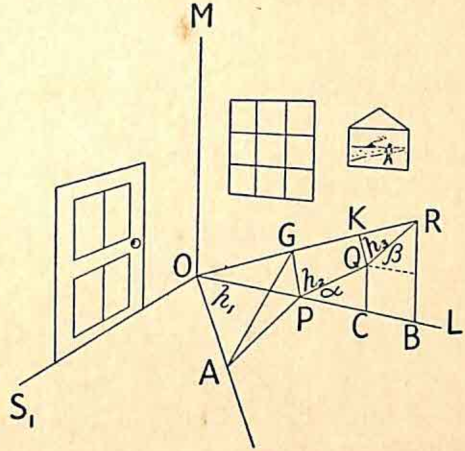

Fig. 12

Since the test-line is in the same plane as h_1 and its own specific, it lies on the floor of the 'room', and OA is unit distance along it. $OP = h_1$. From P draw $PQ = h_2$ and parallel to h_2, and then from Q draw $QR = h_3$ and parallel to h_3. OR is then the centroid of the communal components h_1, h_2 and h_3. (If there had been more tests we would have added h_4, h_5, etc., and R would be at the end of the chain.) We have to show that the cosine of the angle AOR, between the test-line and the centroid of the h's, is the same as the loading of test 1 on the first factor obtained by simple summation. Project P and Q on to OR at G and K, and project Q and R on to OL at C and B.

The angle AGO is a right angle (see Section Three, page 19), and OG is therefore the cosine we seek. We have then

$$\cos AOR = OG = h_1 \cos POG = h_1 \frac{OB}{OR}$$

$$= \frac{h_1(h_1 + h_2 \cos \alpha + h_3 \cos \beta)}{OR} = \frac{h_1{}^2 + h_1 h_2 \cos \alpha + h_1 h_3 \cos \beta}{OR}$$

$$= \frac{h_1{}^2 + r_{12} + r_{13}}{OR}$$

The numerator is the sum of the first column in the summation process applied to

$$\begin{matrix} h_1^2 & r_{12} & r_{13} \\ r_{12} & h_2^2 & r_{23} \\ r_{13} & r_{23} & h_3^2 \end{matrix}$$

It remains to show that OR^2 is the sum of all the items in this matrix, thus:

$$\begin{aligned} OR^2 &= OB^2 + BR^2 \\ &= (h_1 + h_2 \cos \alpha + h_3 \cos \beta)^2 + (h_2 \sin \alpha + h_3 \sin \beta)^2 \\ &= h_1^2 + h_2^2 + h_3^2 + 2h_1h_2 \cos \alpha + 2h_1h_3 \cos \beta \\ &\qquad\qquad + 2h_2h_3 \cos (\beta - \alpha) \\ &= h_1^2 + h_2^2 + h_3^2 + 2r_{12} + 2r_{13} + 2r_{23} \end{aligned}$$

Thus the loading of the first test with the first simple summation factor is the cosine of the angle between the test-line and the centroid of the communality components, and the same for each other test.

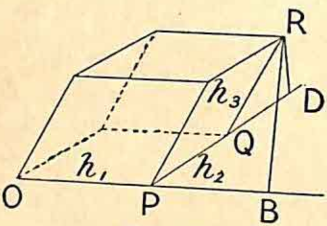

FIG. 13

The above proof is for a common-factor-space of two dimensions: but the same form of proof serves for any number of c.f.s. dimensions. Figure 13 is for three common factors. The cosine required is still the projection of h_1 on to OR the centroid (when the c.f.s. is of three dimensions the test-line cannot appear on the paper under ordinary perspective, being in a higher dimension). As before,

$$OB = h_1 + h_2 \cos \alpha + h_3 \cos \beta \ (+ \ . \ . \ . \text{ if more tests})$$

and OR^2 will still be found by the reader (to whom detailed working is left) to equal the sum of all the items in the correlation matrix with communalities in the diagonal.

SECTION EIGHT

MULTIPLE CORRELATION AND PARTIAL REGRESSION COEFFICIENTS

WHEN we know the correlation coefficients between a number of variates $x_0, x_1, x_2 \ldots x_n$ taken in all pairs, it is possible to find weights for the variates x_1 to x_n so that the correlation between their weighted sum and the other variate x_0 is a maximum. This maximum correlation is commonly called the multiple correlation, and the weights, partial regression coefficients. (We shall assume the variates to be standardised. If they are not, the coefficients we speak of will require each to be divided by the standard deviation of its variate.)

When the variates are the scores in mental tests, x_0 is usually something the experimenter wants to predict, such as secondary school success predicted from a battery of tests given before entrance to the school. It is often referred to by psychologists as the 'criterion', or sometimes as the 'predicand'.

Figure 14 portrays the case of two tests x_1 and x_2 and one predicand x_0. The lines for x_1 and x_2 are drawn on a table-top to assist in visualising the fact that the line for x_0 is not in their plane. A weighted combination of x_1 and x_2 is to be found which will be as near as possible to x_0—for the smaller the angle with x_0 the larger the cosine and therefore the larger the correlation. Now no combination of x_1 and x_2 can be elsewhere than on the table-top: and clearly the line on the table-top which is nearest to x_0 or OD is the line OK, the projection of OD on to the table. If the distance OD is unity, we shall show that the proper weights to give to x_1 and x_2 are the distances OA and OB, obtained by drawing the parallelogram KAOB. The parallelogram of forces show us this. It is necessary then to show that these distances (OD being unity)

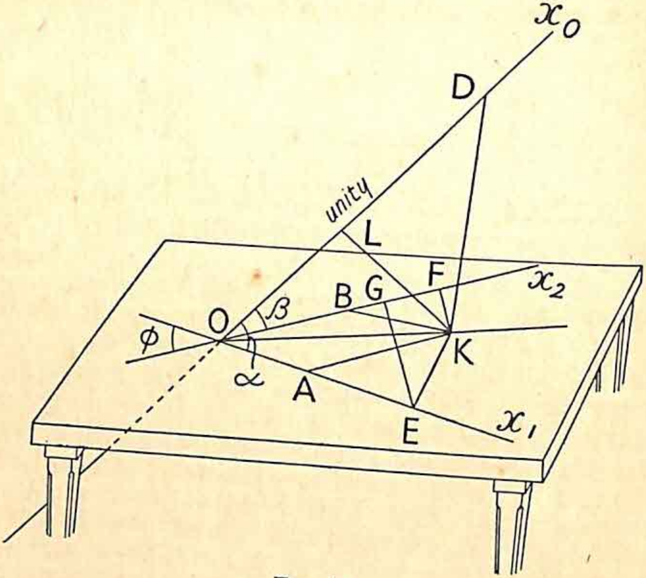

FIG. 14

are the same fractions as are given by the formulæ for partial regression coefficients found by the method of least squares. For two tests these are

$$\left.\begin{aligned} b_1 &= \frac{r_{01} - r_{02}r_{12}}{1 - r^2_{12}} \\ b_2 &= \frac{r_{02} - r_{01}r_{12}}{1 - r^2_{12}} \end{aligned}\right\} \qquad (10)$$

To show this, in Figure 14 drop perpendiculars KE and KF to the two test-lines, and from E drop a perpendicular EG to OF. Then (by Section Three, page 19) the angles DEO and DFO are right angles, and

$$\text{OE} = \cos \alpha = r_{01}$$
$$\text{OF} = \cos \beta = r_{02}$$

for the correlation coefficients are represented by the cosines

	x_0	x_1	x_2
x_0	1	$\cos \alpha$	$\cos \beta$
x_1	$\cos \alpha$	1	$\cos \phi$
x_2	$\cos \beta$	$\cos \phi$	1

Further, $\quad \text{OG} = \text{OE} \cos \phi = \cos \alpha \cos \phi$

so that $\quad \text{GF} = \text{OF} - \text{OG} = \cos \beta - \cos \alpha \cos \phi$

But GF is the projection of KE and equals KE $\sin \phi$, and KE is the projection of OB and equals OB $\sin \phi$, so that

$$\text{OB} = \frac{\text{KE}}{\sin \phi} = \frac{\text{GF}}{\sin^2 \phi} = \frac{\cos \beta - \cos \alpha \cos \phi}{1 - \cos^2 \phi} = \frac{r_{02} - r_{01} r_{12}}{1 - r^2_{12}}$$

Similarly OA is the other partial regression coefficient. Thus the weights to be given to scores in two standardised variates (*e.g.* two tests) so that their weighted sum will correlate as highly as possible with a third variate, are equal to the oblique co-ordinates, on the two test-lines, of a point projected from unit distance along that third variate on to their plane.

This statement is also true for a battery of more tests, and a criterion.

In our simpler diagram, Figure 14, the correlation between the variate x_0 and the weighted sum of x_1 and x_2 is the cosine of DOK, which equals OK since OD is unity. That is, the multiple correlation is measured by OK. If we project OK on to x_0, at OL, then OL measures the square of the multiple correlation. Now OL is also the projection of the chain OA and AK on to x_0, that is, it is equal to

$$\text{OA} \cos \alpha + \text{AK} \cos \beta$$

or

$$b_1 r_{01} + b_2 r_{02}$$

if we call the partial regression coefficients b_1 and b_2. That is,

$$r^2_{\max} = b_1 r_{01} + b_2 r_{02}$$

where $r_{\max}$ means the multiple correlation. It is not difficult

to see that such an equation will hold for more variates in the battery. For if K is the projection of D on to a battery *space*, and OK is projected on to the criterion line as OL, OL will still measure the square of the cosine of the angle DOK. Also OL will still equal the sum of the projections of the oblique co-ordinates of K on to the criterion line, *i.e.*

$$r^2_{\max} = b_1 r_{01} + b_2 r_{02} + b_3 r_{03} + \ldots \tag{11}$$

* * * * * * *

The remainder of this section is for readers with more mathematical knowledge, namely a knowledge of the method of least squares and the elements of matrix algebra. Other readers may nevertheless wish to read the geometrical discussion of Figure 16 (pages 45–47).

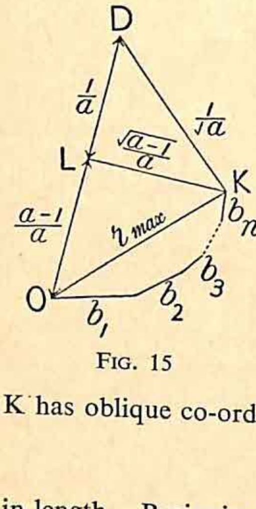

FIG. 15

Figure 15 is intended to represent $n+1$ dimensions, for a battery of n tests and a criterion, unit distance along which is represented by OD. When OD is projected on to the n-space of the tests we get OK, and OK is projected back on to the criterion line giving OL. The point K has oblique co-ordinates along the test-lines measuring

$$b_1, b_2, b_3 \ldots b_n$$

in length. Beginning at O, a chain of lines equal to these b's in length, and parallel to them in direction, is built up. The chain will end at K.

Let the correlation coefficients of the tests and criterion with each other be

$$\begin{bmatrix} 1 & r_{01} & r_{02} \cdots r_{0n} \\ r_{01} & 1 & r_{12} \cdots r_{1n} \\ r_{02} & r_{12} & 1 \cdots \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ r_{0n} & r_{1n} & \end{bmatrix} = \mathrm{R}$$

and let $|\mathrm{R}| = \Delta$.

Let these correlations be represented by

$$\begin{bmatrix} 1 & \cos\alpha & \cos\beta & \cos\gamma \cdots \\ \cos\alpha & 1 & \cos\phi & \cos\theta \cdots \\ \cos\beta & \cos\phi & 1 & \cos\psi \cdots \\ \cos\gamma & \cos\theta & \cos\psi & 1 \quad \cdots \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

Further, let a be the leading entry in the reciprocal of the matrix R, *i.e.*

$$a = \frac{\Delta_{00}}{\Delta} \tag{12}$$

where Δ_{00} is the minor of the first item in R. Then it can be shown that the distances OL, LD, LK and KD in Figure 15 have the values there shown in terms of a.

We can now show that geometrical considerations in the figure lead to values of $b_1, b_2, b_3 \ldots$ etc. identical with those found by the method of least squares algebraically. We shall write the work out for $n = 3$ only, but it is clearly quite general.

The projection of OK on to the line of test 1 is the same as the sum of the projections of its chain of b's. It is also,

by the principle of Section Three (page 19), equal to the direct projection of OD on to that line. That is,

$$\cos \alpha = b_1 + b_2 \cos \phi + b_3 \cos \theta$$

Similarly $$\cos \beta = b_1 \cos \phi + b_2 + b_3 \cos \psi$$

and $$\cos \gamma = b_1 \cos \theta + b_2 \cos \psi + b_3$$

In matrix form,

$$\begin{bmatrix} \cos \alpha \\ \cos \beta \\ \cos \gamma \end{bmatrix} = \begin{bmatrix} 1 & \cos \phi & \cos \theta \\ \cos \phi & 1 & \cos \psi \\ \cos \theta & \cos \psi & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

whence $$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} \text{reciprocal of} \\ \text{the above} \\ \text{matrix} \end{bmatrix} \begin{bmatrix} \cos \alpha \\ \cos \beta \\ \cos \gamma \end{bmatrix} \qquad (13)$$

But this is exactly the same solution for b_1, b_2 and b_3 as is found algebraically by the method of least squares. We wish by that method to find values for b_1, etc. which will, as nearly as possible, ensure that the weighted sum of their scores in three tests will, for each of N persons, give values identical with their scores w_i in the criterion. This is impossible with only three (or $n < \mathrm{N}$) constants at disposal in N equations. The best that can be done is to ensure that the difference between the criterion score and the weighted battery score is small in each case, say v_i. We then have N equations

$$w_1 - (b_1 x_1 + b_2 y_1 + b_3 z_1) = v_1 \quad \text{for the first person}$$

$$w_2 - (b_1 x_2 + b_2 y_2 + b_3 z_2) = v_2 \quad \text{for the second person} \qquad (14)$$

and so on for N persons. By the principle of least squares we minimise Σv^2. This leads to the three 'normal equations'

$$\Sigma wx - b_1 \Sigma x^2 - b_2 \Sigma xy - b_3 \Sigma xz = 0$$

$$\Sigma wy - b_1 \Sigma xy - b_2 \Sigma y^2 - b_3 \Sigma yz = 0$$

$$\Sigma wz - b_1 \Sigma xz - b_2 \Sigma yz - b_3 \Sigma z^2 = 0$$

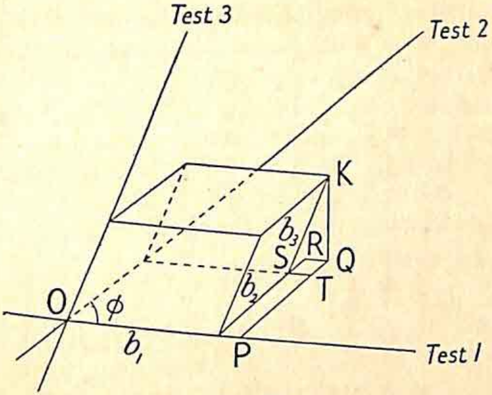

FIG. 16

and these equations are equivalent to

$$r_{01} - b_1 - b_2 r_{12} - b_3 r_{13} = 0$$
$$r_{02} - b_1 r_{12} - b_2 - b_3 r_{23} = 0$$
$$r_{03} - b_1 r_{13} - b_2 r_{23} - b_3 = 0$$

that is, to

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 1 & r_{12} & r_{13} \\ r_{12} & 1 & r_{23} \\ r_{13} & r_{23} & 1 \end{bmatrix}^{-1} \begin{bmatrix} r_{01} \\ r_{02} \\ r_{03} \end{bmatrix}$$

exactly the same equations as those arrived at by the geometrical argument, if we write cosines for correlation coefficients. We have only written this out for a battery of three tests, but it is clearly general for n tests or variates.

The model also enables us to see just what happens when a variate is removed from a battery, resulting in a decrease in the multiple correlation of the remaining battery with the criterion and a change in the partial regression coefficients. Figure 16 is for three tests, and the criterion line, not shown, is in a fourth dimension. As in other diagrams, OK is the projection on the test-space of unit distance OD along the criterion, and equals the multiple correlation. Its oblique

co-ordinates along the test-lines are the partial regression coefficients b_1, b_2, b_3. What happens to these if test 3 is removed from the battery, leaving only tests 1 and 2? In that case, OD on the criterion line would have to be projected on to the plane of these two tests at Q, and the new multiple correlation would be measured by OQ. But, by the principle of Section Three (page 19), Q is also the projection of K on to the plane. So the reduction in the square of the multiple correlation is

$$OK^2 - OQ^2 = KQ^2$$

and the new partial regression coefficients are OP and PQ. The increases in the coefficients b_1 and b_2 are therefore ST and TQ = SR. Now the figure KSTQR is clearly a miniature of a figure like our Figure 14, for finding the weights to give to tests 1 and 2 in estimating test 3; and because SK $= b_3$ instead of unity, the lengths ST and SR are b_3 times these weights. That is, when a test i is omitted from a battery, the partial correlation coefficients of the surviving tests have to be increased by b_i times their coefficients in estimating the omitted test, a statement true for a battery of many tests, not merely of three, as imagining a Figure 16 in many dimensions will show.

If the intercorrelations of the tests of the battery are r's, and the elements of the reciprocal matrix are c's, then these increases are, for the abolition of test i,

$$-\frac{c_{ij}}{c_{ii}}b_i$$

where b_i is the partial regression coefficient of test i in the battery before its omission. That is, the coefficient of each test j is increased from b_j to

$$b_j - \frac{c_{ij}}{c_{ii}}b_i$$

by the omission of test i. (The quantity c_{ij} is commonly, though not necessarily, negative.)

The reduction in $r^2_{\max}$ is in Figure 16 KQ^2, and

$$\begin{aligned} KQ^2 &= KS^2 - SQ^2 \\ &= b_3{}^2 - (ST^2 + TQ^2 + 2ST\,.\,TQ \cos\phi) \end{aligned}$$

and a little algebraic manipulation, using the formulæ in b's and c's for ST and TQ, and remembering that $\cos\phi = r_{12}$, reduces this to

$$\text{decrease in } r^2_{\max} = \frac{b_3{}^2}{c_{33}}$$

or, in general,
$$\frac{b_i{}^2}{c_{ii}}$$

SECTION NINE

PREDICTION OF A MEASURABLE VARIATE

LET us suppose, in the first place, that the scores of a very large number N of persons are known in $n+1$ variates (or tests), that the correlation coefficients have all been calculated; but that the list of scores in the first test x_0 has been lost. Can we reconstitute that list from a knowledge of each person's n scores in the other tests, and of the correlation coefficients?

The answer is that we cannot say what each of the N persons scored in the missing list. But for each group of them all with the same set of scores in the n surviving tests, we can say what the average score of the group was in the missing test list. If we now give to each member of this group that average score of his group in x_0, then this will in half the cases be too big and in half too little, but it is the best we can do. We can also, by adding a 'plus or minus' quantity, indicate how the actual scores in the recovered list will be found to be scattered about this 'prediction'.

In this situation of a lost list 'prediction' is hardly the right word: but in practice actual prediction is usually required and is effected—though with still less certainty—by the same procedure. Imagine that all the correlation coefficients of the $n+1$ tests have been calculated from scores made by one set of persons, and that another set of persons is tested in the n tests but not in x_0. We ask what they will score in x_0 when (perhaps much later) they come to take that test. Each of them can certainly be allotted a predicted or estimated score in x_0, but it will only be the average of what he and others just like him in the n tests will later score, it will not be exactly correct for him individually. And here, where two different sets or samples of people are concerned, there will be an additional error unless the samples are exactly similar.

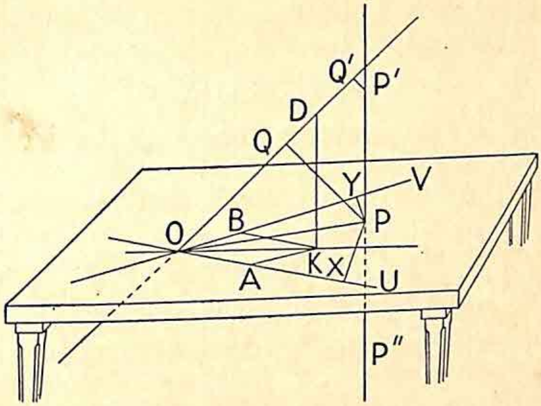

FIG. 17

We shall consider, geometrically, the case where $n=2$, and point out that the proof is general for a battery of n tests. In Figure 17 test-lines are shown for two tests, OU and OV, and a 'criterion' line OD for x_0 which has to be 'predicted' from scores in the two tests. The space is three dimensional, and the points representing persons are spherical in contour round the point O where the three test-lines cross. P represents a person who gets scores OX and OY in the two tests—but note that every person whose point in the three-space is vertically above or vertically below P also gets these same two scores. From the two scores we cannot tell where a person's actual point is on that vertical line. P is at the mid-point of their distribution along it, and to find the average score the group will make in x_0 we project P on to its line at Q, and then OQ is the score in question.

The algebraic formula for the 'best' prediction of a score in x_0 from scores in x_1 and x_2 (best in the least squares sense) is

$$\hat{x}_0 = b_1 x_1 + b_2 x_2$$

(compare equation (14), page 44), where the circumflex accent over x_0 indicates that it is not a measured but only a predicted

value. If our model agrees with this, then we ought to be able to show that (when OD is unity)

$$\mathrm{OQ} = \mathrm{OA} \times \mathrm{OX} + \mathrm{OB} \times \mathrm{OY}$$

Call the angles UOV and UOP, ϕ and ω. Then the right-hand side is

$$\begin{aligned}
&\mathrm{OA}.\mathrm{OP} \cos \omega + \mathrm{OB}.\mathrm{OP} \cos (\phi - \omega) \\
&= \mathrm{OP}\{\mathrm{OA} \cos \omega + \mathrm{AK} \cos (\phi - \omega)\} \\
&= \mathrm{OP} \times \text{projection of OK on to OP} \\
&= \mathrm{OP} \times \text{projection of OD on to OP} \\
&= \mathrm{OP} \times \cos \mathrm{DOP} \quad \text{(since OD is unity)} \\
&= \mathrm{OQ}
\end{aligned}$$

Moreover, this proof is clearly general for any number of tests, *i.e.* any number of dimensions in the battery space instead of the two dimensions of the table-top. The point K is always the projection of D on to the test-space, and the oblique coordinates of K along the test-lines are the *b*'s, the partial regression coefficients. The scores (like OX) are OP times a cosine. The cosine is transferred to the coefficient (like OA cos ω), and these add up to OK projected on to OP, *i.e.* to cos DOP. So the value of the whole is OP cos DOP or OQ.

We have in this section been considering the recovery of a lost set of test-scores in one test from a knowledge of the scores in a battery of other tests with known correlations with the missing one and with each other. When the lost list is found, we can compare our predictions or estimations with it, and we find that we are only right in an average way. Half our estimates are too high, half too low.

We have applied the same procedure to build up a list which has been not 'lost' but not yet made. Again, when it is made, we can compare our predictions with the facts, and find this time that even our averages for similar subgroups may be wrong, since the correlation coefficients must have been calculated from a different sample of persons.

In spite of these uncertainties, the method is the best we can do when the practical need to predict is great, and we can

temper assurance with 'plus or minus' added to each prediction. This range of uncertainty will be the narrower, the nearer the predicand or criterion line is to the test-space, *e.g.* in Figure 17, the smaller the angle DOK: or perhaps we should say, the nearer the test-space is to the predicand line, for it is the former we can change, by picking suitable tests. When the predicand line is actually in the test-space we can predict with certainty.

SECTION TEN

THE ESTIMATION OF FACTORS

A FACTOR is a hypothetical test, and its correlation coefficients with the actual tests of a battery are its so-called loadings or saturations in each test. These, for centroid factors, are found by the process of simple summation: and other factors can be found by rotating the centroid factors to new positions.

The situation, then, if we want to estimate a man's factors from his scores in a battery of tests, is exactly the same as when we want to estimate his scores in a further test, except indeed for the important difference that in the latter case we can give him the criterion test and check our estimate. We cannot give him the hypothetical factor test. It doesn't exist.

When communalities are used, none of the factors, common or specific, is in the test-space. The method described in the preceding section can be used to estimate a man P's factors in exactly the same way as is there described for estimating or predicting his score in a real test. In Figure 17, OD could be a factor-line. Factor-estimations made in this way have the same advantages and disadvantages as estimates of a real test-score from correlated tests. They will give every one of a group of men who get the same test-scores, the same factor-estimates, although their factors no doubt differ round this as mean in the same way as with the real criterion.

A factor can be considered as the resultant of two components, one in the test-space and one at right angles to it. The scores a man makes in the tests can tell us nothing whatever about this second component, which is absolutely uncorrelated with any of the tests we have given him. All we can do is to make some assumption about his ability in that direction, and the assumption made when we apply to a factor the method of estimation described in the preceding section is that in this unknown ability at right angles to the test-space he is average.

We take P (Figure 17) in the test-space as being actually his point also in the space of higher dimensions. It is important to stress the fact that when there are more factors than tests (as there are when communalities are used) nothing whatever, no mathematical device whatever, can do away with the need for making some such assumption. And the assumption of average ability, when ignorance is complete, is the proper one to make—we shall call this 'assumption A' (two others will be spoken of later).

Someone may protest against this, saying that if P does well in the battery of tests he is likely to be above average in any new test, for abilities tend to be positively correlated. True; but not this particular ability, for it is at right angles to all the battery tests, uncorrelated with any of them. We have no reason at all to think that P is good or bad in this direction, in this component of his factor. Since more men are average than anything else, we had better assume P to be average.

Method A, then, uses to assess a man's factors exactly the same procedure as is commonly used in assessing a real and directly measurable quantity, the method described in the preceding section.

Method B, proposed by Professor M. S. Bartlett, makes a different approach, by minimising the influence of the specific factors. His working out of this is done algebraically, but it can, at any rate crudely, and perhaps exactly, be said to be equivalent to assuming, not that P is average in the unknown component, but that he has such ability in it as places his point as near as possible to the common-factor-space. In the simple battery of two tests portrayed in Figure 17, this means that P is moved to P′, where the vertical line most closely approaches the criterion line OD: and then Q′, its projection on to OD, gives the value OQ′ for the criterion estimate. (In the solid Figure 17, the vertical line PP′ passes behind the continuation of OD: for OD is vertically above OK.)

When there are several common factors, and a c.f.s. of several dimensions, the problem is to find a position for P′ as near as possible to that common-factor-space, while still keeping P′P at right angles to the test-space.

For estimating a real measurable variate this procedure is quite inadmissible. For estimating factors, it has an advantage in the eyes of those who consider that it is improper to maximise the specifics, as is done by the common practice of minimising the communalities: for this Method B minimises them again at a later stage. But note that by doing so it has changed the factors. If smaller specifics, that is larger communalities, had been used at the outset, the common factors would have been different and there would have been more of them.

Before discussing the third assumption made about each man's ability at right angles to the test-space, it is necessary to devote a section to the consideration of the presence of correlation among factor-estimates even of uncorrelated factors when the estimates are made by either Method A or by B.

SECTION ELEVEN

THE CORRELATION OF FACTOR-ESTIMATES

THE factors we have had in mind throughout have been orthogonal, that is, uncorrelated factors. Their lines in our model are at right angles to each other and form a rectangular co-ordinate system of the factor-space. Centroid factors by their method of calculation, and specific factors by their very definition, are orthogonal to one another, and so too are most systems of factors derived from these by rotation. (Some modern systems use correlated factors, but these we have not so far mentioned.)

Yet when these uncorrelated factors are estimated their estimates are correlated! Essentially this is due to the fact that while the points representing persons are, in Method A, left in the test-space, the factor-lines are outside it. (The person-points in Method B are moved out of the test-space, and it seems probable that Bartlett estimates are less correlated than those made by Method A.)

In Section One it was emphasised that, in the original scattergram, correlation between variates represented by lines at right angles was shown by the ellipticity of the contours of density of the points representing persons—points such that their co-ordinates were the test-scores of the respective persons. If, in such a scattergram, the contours were circular, there would be no correlation. For a given score x, the average value of y would be zero, there would be no tendency for it to alter as x altered.

With a solid scattergram for three tests (their lines at right angles to one another), the points representing persons will be spherical in density-contours if there is no correlation present, ellipsoidal if there is. Such scattergrams can be imagined also for more tests, in more dimensions.

In our Section One we changed this by certain compressions and extensions of the swarm of points until they were always spherical whatever the correlations. The test-lines were then no longer orthogonal but at angles whose cosines measured the correlation coefficients. The centroid factors described in Section Six, where full communalities of unity were used, are also in that test-space, and equal in number to the tests. There are no specific factors in that section. The factor-lines are at right angles to one another, the swarm of points is spherical, and there is no correlation between the factors or between their 'estimates', which here indeed are not mere estimates but measures. The factors have no components outside the test-space, and it is immaterial what P's ability is, at right angles to that space.

It is different, however, as soon as fractional communalities are used, and specific factors required. The total number of factors is then greater than the number of tests, and the factor-space of more dimensions than the test-space. In that factor-space, the points representing the N persons whom we have tested are no doubt distributed spherically in density: but we do not know where they are, we only know their projections on the test-space. Those projected points, in that test-space, are spherical in density; but a sphere in n dimensions is, in say $n+c$ dimensions, an ellipsoid with c of its axes zero. So the picture, in the factor-space, is one of ellipsoidally distributed points, and therefore of correlation between the factor-estimates if these points are used, as in Method A, to project on to the factor-lines for each man's estimated factors.

If this explanation is unconvincing to a reader unaccustomed to thinking spatially in high dimensions, a consideration of the simple case portrayed in Figure 9 (page 33) may be illuminating. There only two tests form the battery, and there are three factors, G, S_1 and S_2 whose lines define a three-space. The test-space is a plane, partly shaded in the figure. Only the positive halves of each test-line and each factor-line are shown. The test-lines, for example, really go on below the floor of the 'room' indicated by the factor-lines.

Now the density-contours of points representing persons

on the sloping plane which is the test-space are circles. But when these are projected on to one of the walls of the room, say on to GOS_1, they become ellipses, and the estimates of the factors G and S_1 are therefore correlated. In this there is no difficult spatial imagination demanded, for everyone knows that the shadow of a circular disc on a wall is an ellipse, if the disc is not parallel to the wall and the light perpendicular.

By analogy, the projection of the circular outline of an *n* dimensional sphere on to a plane not in that *n* dimensional space is also an ellipse, and thus correlation is created between the two factor-lines defining that plane. All this is due to using the points representing persons in the test-space as also representing them in the wider factor-space. Their points in that wider factor-space, did we know them, are really distributed spherically in it: but we do not know them, only their projections on the test-space.

SECTION TWELVE

UNCORRELATED FACTOR-ESTIMATES

THE situation is best approached by again considering a battery of two tests. In Figure 17 (page 49) let U and V be the lines of two tests, and the line OD a factor-line. The table-top is the test-space, the factor-space is of more dimensions: for the moment let us think of it as a three-space.

The very numerous points P (N in number) are on the table-top, and are circular in density, round O. Along any line on the table-top the points are distributed normally. At any one position like P in the figure, there will be not one point but quite a number (if N is large), being the projections of all the points situated on the vertical line P′PP″, both above and below the plane.

Now we do not know where these persons, concentrated at P (having all scored OX and OY) really are situated up and down the line P′PP″. But we do know how they are distributed up and down that line normally. We can therefore, if we are ruthless enough, allot to each of them a position, moving them vertically different distances away from the test-space (that is, making different assumptions about their abilities, although in the tests they are identical) until their distribution is normal with unit standard deviation. If we do this for each possible position of P all over the table-top, the N points, instead of being a disc-like crowd on the plane, will be a solid swarm round O, and their contours of density will be spheres. The process is just the opposite of projecting a solid sphere on to its equatorial plane. We are given the distribution over the equatorial plane and required to reconstitute the sphere. We can reconstitute *a* sphere, but except by a miracle the points will not be back where they were before the projection, it will not be *the* sphere.

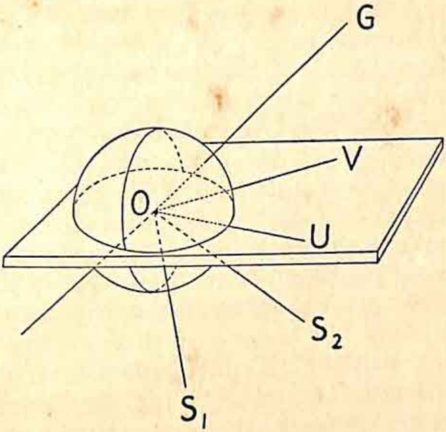

FIG. 18

At this point let us, for clarity, rub out the parts of Figure 17 we are not using (see Figure 18) and add a second and a third factor-line, OS_1 and OS_2. In the figure, one of the circular contour lines on the table is shown expanded up and down into a sphere. Points on the table have been moved up, or down, different distances till their configuration in the three-space of the factors is spherical. When these positions are projected on to OG, OS_1 and OS_2 they will give values for the factor-estimates which are uncorrelated.

But it should be noted that this does not mean we can give factor-estimates to individual men. If Tom, Dick, Harry and others are all at one point on the table-top (that is, have each the same pair of scores in the tests U and V), we know that some of their points in the three-space of the factors must be above the table, and some below; and we can scatter them up and down in the correct distribution; but we do not know whether Tom goes up and Dick down, or *vice versa*. We can arrive at a set of uncorrelated estimates, but do not know which man has to be given which estimate! This makes the procedure useless for any practical purpose, such as vocational advice.

What has been illustrated from the crudely simple case of two

tests and three factors is also true for n tests and $n+c$ factors. The scores of N persons in the n tests give that number of points spherically distributed in the test-space. From the point of view of the $n+c$ space, however, these points are ellipsoidally distributed, and estimates of factors made by projecting them on to the factor-lines are therefore correlated. The N points can of course be moved orthogonally away from the test-space in different directions and by various amounts until their distribution in the factor-space is spherical, and their projections will then give uncorrelated factor-estimates. But we cannot allot these estimates to individual men, so they are not of practical use.

A brilliant paper by H. Kestelman in Vol. V of the *British Journal of Statistical Psychology* has shown, using matrix algebra, how the values of the uncorrelated factor-estimates can be calculated. It does not, of course, mean—though the unwary reader might mistakenly think so—that these values can be given to individual men. That would be quite unjustifiable.

Another way of looking at the matter may make the situation clearer. Suppose, in Figure 17, we had given to each of the several persons at P the estimate OQ. This set of such estimates would be correlated with estimates similarly made of another factor. But we could correct this by altering the estimates OQ until they were scattered, keeping the mean at Q, with the proper standard deviation, which is equal to the cosine of the angle between the factor-line OD and the perpendicular to the test-space, that is to the square root of $(1-r^2_{\max})$. When this is done for every P and its Q, the N estimates along the factor-line will have a standard deviation of unity, and will be uncorrelated with those along other factors.

"THE CRITICAL ISSUES IN PSYCHOLOGY" Serie[illegible] Melvin H. Marx, is designed to provide authoritative an[illegible] tive contemporary reviews of selected, important topics w[illegible] central and germane to a spectrum of courses in psychology. Each [illegible] ume combines an insightful original essay and a set of reprinted reports especially chosen as distinguished instances of fundamental progress or position, and thus presents superbly useful text. Emphasis throughout the series is basically empirical rather than speculative, and the raising of fundamental questions is stressed. Written by active researchers, these publications will offer instructor and student opportunity to read sound, high-quality materials on a variety of topics not adequately covered in ordinary textbook discussions. It is hoped that an important contribution of the series will be the stimulation of interest of the scientifically-oriented student in the intellectual challenges offered in psychology.

The series editor Melvin H. Marx, Professor of Psychology at the University of Missouri, is currently president of the Midwestern Psychological Association (1964-1965). He recently edited "Theories in Contemporary Psychology" (Macmillan); he is consulting editor for Macmillan Publications in Experimental Psychology.

The Macmillan Company, 60 Fifth Avenue, New York 10011

The Critical Issues in Psychology Series
MELVIN H. MARX, General Editor